Arpeggion-H: Uma Interface Interativa para Reprodução de áudio MPEG-H

Bruno Augusto R. de M. Moreira
brunoarmm@id.uff.br
Universidade Federal Fluminense
Niterói, Brasil

Debora C. Muchaluat-Saade
debora@midiacom.uff.br
Universidade Federal Fluminense
Niterói, Brasil

ABSTRACT

This paper provides an overview of the Arpeggion-H audio player, able to reproduce and interpret the metadata extracted from mhm1 encoded MP4 files, based on the MPEG-H part 3 norm. It is also able to display and alter the available configuration inferred from the audio's metadata at any time. This tool will be useful to academic and common users, as it is able to play and allow user interaction out of a codec that is normally not supported. In the present implementation, this tool uses the decoder implement by the Fraunhofer-IIS group called "mpeghdec", but it may support any other decoder if properly integrated.

KEYWORDS

MPEG-H 3D Audio, Interatividade, Audio player, Cena de áudio

1 INTRODUÇÃO

A norma MPEG-H [4] foi criada visando facilitar a criação de áudios 3D com descrições de cenas, nas quais o usuário pode interagir e configurar. É possível, em um arquivo MP4, definir essas descrições da cena de áudio, inseridas pelo codec mhm1. Essas definições dividem as entradas de áudio em objetos, e permite dar o controle ao usuário para a configuração das propriedades do áudio, como o volume, linguagem, ou a coordenada do áudio. Além disso, o MPEG-H permite escolher a combinação de alto-falantes que desejar e o decodificador irá transformar a saída conforme esta opção escolhida pelo usuário. Isso permite que o áudio original seja independente da configuração de quantidade e posição dos canais de saída.

Este trabalho apresenta a ferramenta Arpeggion-H, capaz de reproduzir e interpretar os metadados extraídos dos arquivos de áudio codificados com mhm1, da norma MPEG-H Parte 3. A ferramenta visa apresentar ao usuário as opções de configuração deste áudio, e permitir modificações em tempo real dessas configurações. Esta ferramenta será útil para pesquisas e usuários comuns por ser um tocador de áudio de um decodificador que não é comumente disponível. Esta ferramenta não implementa o decodificador, mas sim, utiliza um já existente [4] e, no futuro, pode permitir a utilização de qualquer implementação.

A ferramenta visa receber um arquivo de áudio MP4 codificado com mhm1, e conseguir construir e apresentar ao usuário uma interface baseada nas propriedades de cena extraídas diretamente do arquivo MP4, possibilitando a mudança de valores das propriedades do áudio em tempo real. A ferramenta também dará a possibilidade de mudanças de configurações globais e, através da configuração de linguagem, adaptar os textos apresentados na interface para tal linguagem, quando possível.

A ferramenta tem como público alvo usuários comuns que possuem interesse em tocar estes áudios, já que as opções disponíveis atualmente se resumem a ferramenta proprietária MPEG-H Authoring Suite, ou a hardwares de streaming para televisão. Ela também se destina a desenvolvedores e pesquisadores interessados em como implementar um player com essas capacidades.

2 TRABALHOS RELACIONADOS

O trabalho [1] propõe que sejam implementados argumentos, os mesmos que a ferramenta Arpeggion-H utiliza, que manipulem a saída de áudio. Esses argumentos são inspirações diretas da proposta do MPEG-H, o que permite ao usuário manipular informações como posição dos objetos, pre-definições, ganho de volume, entre outros. Ambos os trabalhos têm como foco a implementação dessas normas em plataformas de TV.

O trabalho [2] descreve a norma MPEG-H e categoriza ferramentas como o Arpeggion-H como um "Receptor". Já o artigo [3] propõe uma forma de implementar a API que seria disponibilizada para uma ferramenta front-end, e implementa um protótipo front-end para a validação dessa API. O Arpeggion-H implementa este front-end de forma similar em questão de funcionalidades, porém visualmente distinto.

3 FERRAMENTA

A ferramenta Arpeggion-H consiste em 3 módulos independentes: o decodificador, o interpretador da cena e o reprodutor de áudio. E um módulo principal que une todos os outros módulos, a interface. Por uma limitação da ferramenta, a comunicação entre os diferentes módulos e a biblioteca de decodificação MPEG-H se dá por comunicação por meio de arquivos, isso inclui o arquivo MP4 de entrada. Visto isso, ainda não é possível integrar esta ferramenta com um serviço de streaming de MPEG-H. Na Figura 1, é possível ver como todos os módulos interagem entre si. Na figura, os módulos em roxo pertence à ferramenta deste trabalho, já os módulos em verde pertencem ao *mpeghdec*, e o amarelo representa a entrada do usuário. A ferramenta foi implementada na linguagem Python, com o auxílio das bibliotecas Tkinter para a interface e Pyaudio para a reprodução de áudio.

3.1 Integração com o Decodificador

Foi utilizado o decodificador *Fraunhofer MPEG-H decoder (mpeghdec)* [4], e a integração foi feita através do binário compilado do código-fonte disponibilizados pela Fraunhofer-IIS. A interação é vista na Figura 2. Conforme a demanda, a ferramenta irá decodificar trechos

In: XXII Workshop de Ferramentas e Aplicações (WFA 2024). Anais Estendidos do XXX Simpósio Brasileiro de Sistemas Multimídia e Web (WFA'2024). Juiz de Fora/MG, Brazil. Porto Alegre: Brazilian Computer Society, 2024.

ISSN 2596-1683

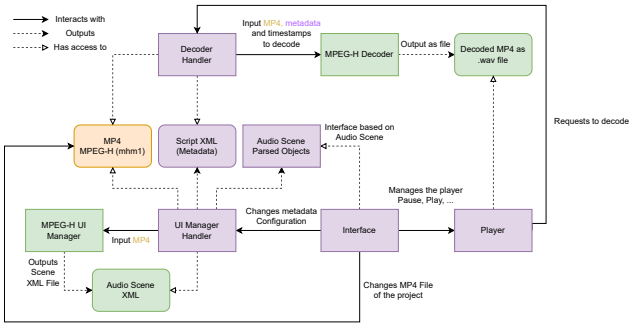

Figure 1: Arquitetura da ferramenta Arpeggion-H, incluindo módulos e a biblioteca *mpeghdec*

pequenos de áudio através do mpeghdec, passando o arquivo a ser decodificado, as alterações que o usuário fizer às configurações, e a marca de tempo a ser decodificada. As alterações de configurações são passadas ao decodificador num formato XML seguindo a especificação das mensagens de ações de evento (*Action Event Messages*).

Além dos argumentos no XML, existem quatro argumentos passados diretamente ao decodificador:

- Composição dos alto-falantes: define o layout. Por exemplo, mono, estéreo ou 7.1;
- Tipo de ambiente: áudio é adaptado para ambientes de diferentes tipos. São eles barulhento, silencioso, noturno, faixa limitada de reprodução, volume baixo, melhora de diálogo e melhor compreensão;
- Coeficiente de ganho (não disponível diretamente na interface);
- Modo Álbum: aplica um fade-in e fade-out entre músicas.

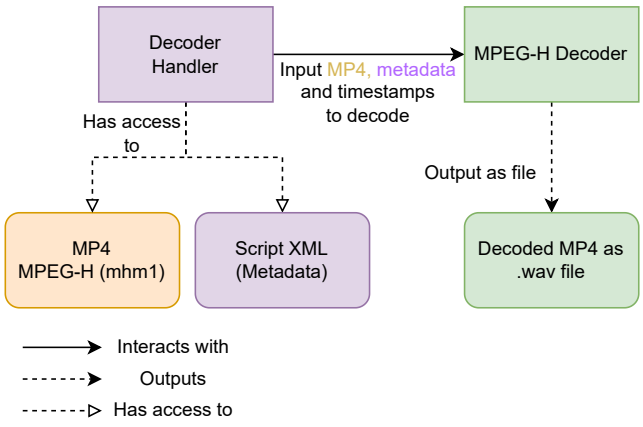

Figure 2: Arquitetura do módulo de decodificação. Roxo pertence à ferramenta deste trabalho, em verde pertence ao *mpeghdec*, e amarelo é entrada do usuário

3.2 Leitor de Cena

O leitor de cena fará o papel de transformar os metadados extraídos do MP4 em algo legível para a interface, como visto na Figura 3. O leitor também utiliza do mpeghdec para gerar um arquivo XML com a descrição de cena de áudio (*Audio Scene Configuration*), e através deste arquivo ele gera essa descrição e disponibiliza para os outros módulos. Além disso, este módulo é responsável em criar o Script XML que contém as alterações do metadado para ser usado durante a decodificação de áudio.

A cena possui um conjunto de pre-definições que o usuário pode escolher, e dentro de cada pre-definição existe um conjunto de áudios e/ou grupos de áudio. Cada áudio e grupo de áudio possui um conjunto de propriedades que o usuário pode ou não alterar, também como seus valores mínimos e máximos. No caso do grupo de áudios, é possível escolher um áudio dentre uma seleção. A definição do formato XML pode ser encontrado em [3].

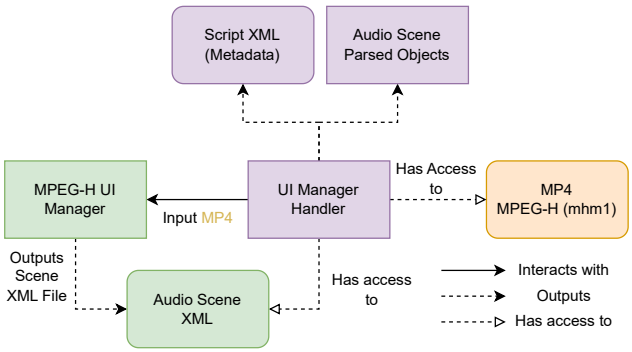

Figure 3: Arquitetura do módulo de Leitor de Cena. Roxo pertence à ferramenta deste trabalho, em verde pertence ao *mpeghdec*, e amarelo é entrada do usuário

3.3 Player

O Audio Player, como visto na Figura 4, utiliza de arquivos em formato wav para reprodução de áudio. Estes arquivos são gerados através do decodificador, que gera esses trechos de áudio. O Player pode ser controlado pela interface e sua implementação é a mais simples possível.

3.4 Interface

A Interface é o componente principal da ferramenta, já que ele integra todas as outras partes, como visto na Figura 5. A Interface utiliza da saída do leitor de cena para construir a interface gráfica visual do usuário, e envia pedidos ao leitor de cena para criar as mensagens de ação de evento baseado na interação do usuário com a interface. A interface também define qual o arquivo MP4 é utilizado pelo resto da ferramenta, e também qual o idioma padrão tanto para os textos quanto para o áudio. A Figura 6 exibe a interface gráfica gerada pela Interface.

É possível ver na Figura 6 como a descrição da cena é representada. Cada pre-definição é uma aba, enquanto cada elemento de áudio é um parágrafo diferente, separados por linhas horizontais cinzas. Todo texto da cena apresentado é extraído diretamente

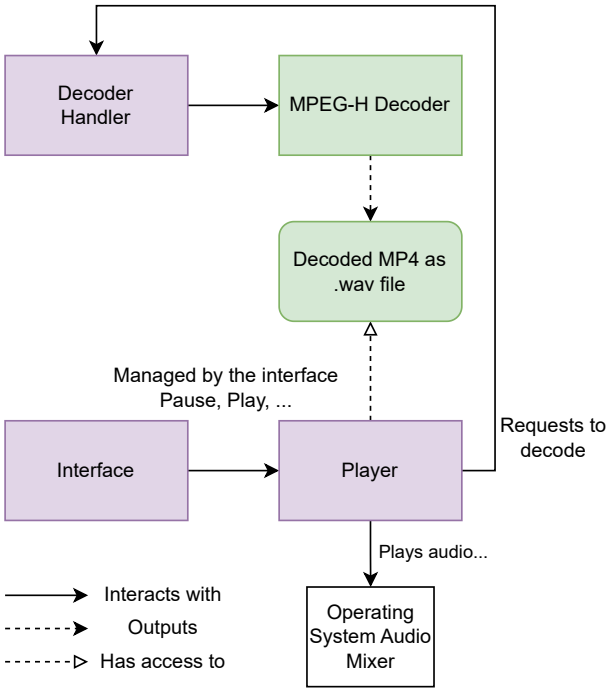

Figure 4: Arquitetura do módulo Audio Player. Roxo pertence à ferramenta deste trabalho, em verde pertence ao *mpeghdec*, e amarelo é entrada do usuário

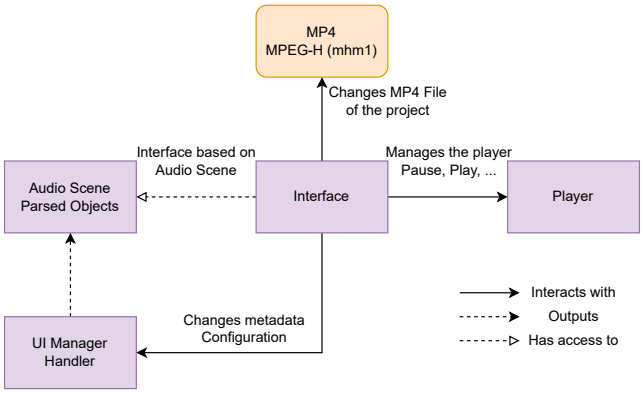

Figure 5: Arquitetura do módulo de Interface. Roxo pertence à ferramenta deste trabalho, em verde pertence ao *mpeghdec*, e amarelo é entrada do usuário

da própria cena. A tradução dos textos depende inteiramente da descrição de cena possuir tais traduções.

4 USO DA FERRAMENTA

Ao abrir a ferramenta, o usuário pode acessar a opção de *File* no topo esquerdo e selecionar qualquer arquivo MP4 que tenha sido codificado com mhm1. Após isso, a ferramenta ira atualizar a interface baseado no áudio recebido e haverá a opção de iniciar a

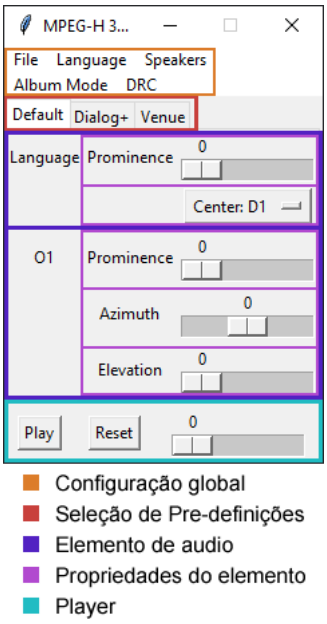

Figure 6: Resultado da leitura da configuração de cena, com algumas partes em destaque

reprodução do áudio através da opção de *Play*. O usuário poderá trocar de arquivo a qualquer momento.

A interface será atualizada ao carregar um novo arquivo, sendo possível visualizar os diferentes elementos de áudio presentes no MP4, onde seus nomes serão mostrados mais à esquerda. Na Figura 7 é possível ver esses elementos de áudio chamados "*Language*" e "*O1*". Ao lado dos elementos, estão presentes os atributos referentes aquele elemento de áudio em específico. As possíveis propriedades são: Prominência (*Prominence*), Rotação (*Azimuth*), Elevação (*Elevation*), Mudo (Muting, não presente na Figura). No caso de um elemento composto de áudio, é possível escolher um dentre um conjunto de áudios, como visto em "Language", a opção "Center: D1" é menu em cascata com outras opções.

O usuário pode interagir com as abas de pre-definições, neste exemplo, chamadas de "Default", "Dialog+", e "Venue".Cada pre-definição altera de alguma forma os valores ou disponibilidade de cada áudio, influenciando no resultado ouvido pelo usuário.

Nas Figuras 8 e 9, é possível ver dois outros resultados de interface de MP4s diferentes.

Como visto anteriormente na Figura 6, existem cinco opções de cascata no cabeçalho da interface: File, Language, Speakers, Album Mode, e DRC. File permite escolher um arquivo diferente, enquanto Language permite escolher a linguagem padrão do usuário. As outras três opções são passadas diretamente ao decodificador como visto na Seção 3.1, sendo elas "Composição dos alto-falantes", "Modo Álbum", e "Tipo de ambiente", respectivamente. Podem ser vistas, na Figura 10, todas as opções disponíveis até o momento pela interface. Serão disponibilizadas mais opções no futuro.

Aplica-se a Arpeggion-H a licença GPLv3. Um vídeo demonstrando a ferramenta pode ser visto em https://drive.google.com/file/d/1_q7Gg37E6WmNuF-a6H_21qvAc4NootGP/view?usp=sharing.

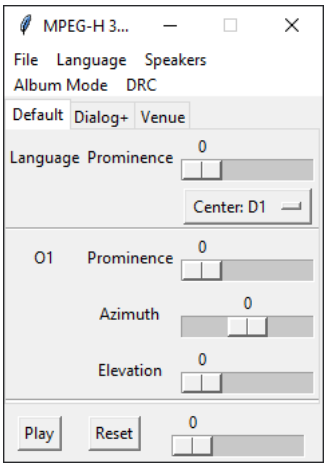

Figure 7: Resultado da leitura da configuração de cena.

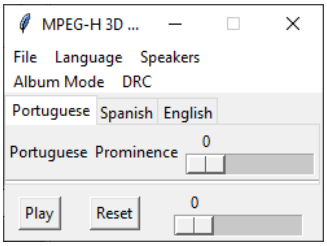

Figure 8: Resultado da leitura da configuração de cena do segundo exemplo

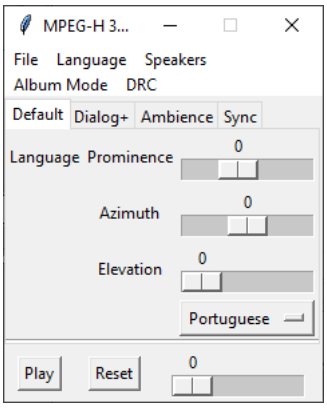

Figure 9: Resultado da leitura da configuração de cena do terceiro exemplo

O código-fonte pode ser visto em https://github.com/BunnyMerz/Arpeggion-H

5 CONCLUSÃO

A ferramenta Arpeggion-H, proposta neste trabalho, consegue reproduzir áudios seguindo a norma MPEG-H, parte 3, e permite ao

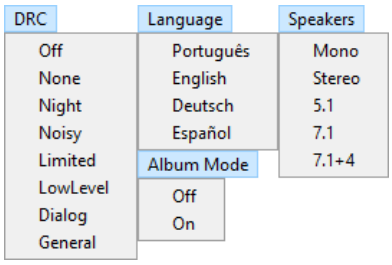

Figure 10: Resultado da leitura da configuração de cena do terceiro exemplo

usuário interagir com as configurações da cena para modificar o áudio reproduzido. A ferramenta adapta a interface gráfica baseado no arquivo de áudio fornecido e apresenta o texto conforme a preferência de linguagem do usuário. Mesmo sendo uma ferramenta simples, será útil para fins acadêmicos, pois reprodutores de mídia deste tipo normalmente requerem licenças pagas e não são de código aberto.

Como trabalho futuro, a ferramenta será estendida para poder receber um streaming de áudio. Para adaptação ao streaming de áudio, seria necessário utilizar e adaptar o código-fonte da Fraunhofer-IIS, já que o binário da compilação padrão dá acesso apenas à decodificação por meio de arquivos. Além disso, estão planejados testes de usabilidade da ferramenta por meio de experimentos com usuários.

REFERENCES

[1] Rafael Diniz and Marcelo F. Moreno. 2019. Immersive audio properties for NCL media elements. In *Anais Estendidos do XXV Simpósio Brasileiro de Sistemas Multimídia e Web* (Florianópolis). SBC, Porto Alegre, RS, Brasil, 195–197. https://doi.org/10.5753/webmedia_estendido.2019.8164

[2] Yannik Grewe, Adrian Murtaza, and Stefan Meltzer. 2023. MPEG-H Audio System for SBTVD TV 3.0 Call for Proposals. *SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING* 7 (Mar. 2023), 17. https://revistas.set.org.br/ijbe/article/view/219

[3] Yannik Grewe, Adrian Murtaza, and Stefan Meltzer. 2023. MPEG-H Audio System for SBTVD TV 3.0 Call for Proposals. *SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING* 7 (Mar. 2023), 10. https://revistas.set.org.br/ijbe/article/view/219

[4] Jürgen Herre, Johannes Hilpert, Achim Kuntz, and Jan Plogsties. 2015. MPEG-H 3D Audio—The New Standard for Coding of Immersive Spatial Audio. *IEEE Journal of Selected Topics in Signal Processing* 9, 5 (2015), 770–779. https://doi.org/10.1109/JSTSP.2015.2411578